How to discover High-quality Roles? A Survey and Dependency Analysis of Quality Criteria in Role Mining

Michael Kunz[1], Ludwig Fuchs[2], Michael Netter[2], and Günther Pernul[1]

[1] University of Regensburg, Department of Information Systems
{michael.kunz,guenther.pernul}@ur.de
http://www-ifs.uni-regensburg.de
[2] Nexis Gmbh, Regensburg
{ludwig.fuchs,michael.netter}@nexis-secure.de
http://www.nexis-secure.de

Abstract. Keywords: Role Quality, Role Mining, RBAC, Identity Management

Abstract. Roles have evolved into the de facto standard for access control in Enterprise Identity Management. However, companies struggle to develop and maintain a role-based access control state. For the initial role deployment, role mining is widely used. Due to the high number and complexity of available role mining algorithms, companies fail to perceive which is selected best according to their needs. Furthermore, requirements on the composition of roles such as reduction of administration cost are to be taken into account in role development. In order to give them guidance, in this paper we aggregate existing role mining approaches and classify them. For consideration of individual prerequisites we extract quality criteria that should be met. Later on, we discuss interdependencies between the criteria to help role developers avoid unwanted side-effects and produce RBAC states that are tailored to their preferences.

Keywords: Role Quality, Role Mining, RBAC, Identity Management

1 INTRODUCTION

Regulating access to resources is an elementary function of every Identity Management System (IdMS). Not just as a result of governmental regulations or compliance requirements like the Sarbanes-Oxley Act [47], Basel III [2], or the EU General Data Protection Regulation in its revised form [13], especially medium- and large-sized companies are forced to control access to sensitive information. Over the past decades, Role-Based Access Control (RBAC [45]), has become the de facto standard for managing access to resources in IT systems. In RBAC, roles act as intermediary between users and permissions, essentially reducing access control complexity. Despite being widely used, RBAC struggles with the dynamic evolution of role models over time. Besides the daily user administration, the central challenge after setting up a role model is its strategic maintenance. Role system maintenance focuses on updating and cleansing role configurations, discarding unused, and defining new roles. Changing business processes, organizational structures, employee positions, or security policies and newly imposed regulations force the administrators to quickly react and adapt the access control structures

Fig. 1. Research methodology

in place. Commonly, this leads to an increasing number of roles, an overall reduction of the role model quality and the advent of security vulnerabilities due to erroneously assigned roles or outdated role definitions.

In order to mitigate the risk of increasing security vulnerabilities in RBAC, one cornerstone of ensuring a high role model quality is the periodic assessment of the role model components, such as the user-role assignments (*UA*), the role-permission assignments (*RA*), or role hierarchy structures. Role mining approaches that support organizations during their initial setup of an RBAC model have attracted the attention of researchers in the last decade. Over the last four years, for instance, a variety of research groups have published approaches to come up with an initial set of roles. However, despite the fact that reports like the Ponemon Cyber Crime Study 2013[3] emphasize the importance of implementing strategic policies and procedures for controlling access control structures, hardly any attention has been drawn to the challenge of maintaining an existing role model. Recently, the need for investigating and cleaning role model structures has been highlighted by [18]. However, the core challenge of measuring the current quality of a role model and select criteria for its optimization still remains unsolved. Due to the development of the area and its importance for access control it is likely that role mining approaches are re-applied by organizations periodically in order to ensure role model correctness. Our work builds on both, the in-depth investigation of the research area as well as our practical experience from several industry projects within medium- and large sized companies dealing with the setup and management of a role-based IdMS. Similar to the work presented in [18], we argue that the practical project requirements cannot be considered to a sufficient extent by the available role mining approaches. In particular, we address the following two research questions:

- *RQ1: Which quality criteria are employed in existing role mining approaches?*
- *RQ2: Which dependencies between those quality criteria do exist?*

Hence, the contribution of this work is threefold. In order to close the existing research gap, this paper firstly analyzes the development of role mining, presenting a short survey of the field and underlining the rapid development in the area. In this respect, it builds on an existing survey of the area published in the year 2011. During execution, various role mining algorithms rely on some sort of quality criteria to different extents. As a result, this paper secondly analyzes and extracts potential criteria for rating the quality of role models included in current role mining approaches. It points out the differences among the various approaches and underlines the need for a structured quality rating process. Thirdly, this paper focuses on the mutual dependencies between the different quality criteria. With this contribution, we aim at stimulating the research

[3] http://www.hpenterprisesecurity.com/ponemon-2013-cost-of-cyber-crime-study-reports

community and engage them to enrich existing role mining approaches considering quality criteria in a structured manner.

The remainder of the paper is structured as follows: In Section 2 we present the RBAC concept and related work. Section 3 outlines the applied methodology before Section 4 shows the conducted survey and categorizes role mining approaches according to their underlying techniques. Subsequently, in Section 5 quality criteria are extracted from existing role detection mechanisms. We discuss their dependencies and show which criteria are mutually exclusive, which are complementary, and which have no effect on others. Finally, Section 6 concludes the paper and outlines future work.

2 RELATED WORK

In todays medium and large-sized companies, RBAC has become the state-of-the-art standard for controlling user access to resources. As a result of the large amount of research output, several surveys of the general area of roles in IT security have been presented (e.g. [64] and [21]). Authors lately agreed upon the growing importance of role development in general and automated role mining in specific. Fuchs et al., for instance, provided an evaluation of role development approaches in [20] and [19]. Since their publications, the research output in the area has grown more than double, requiring a survey update in order to give an in-depth understanding of recent developments.

During the initial setup of a role model, role mining algorithms inherently rely on different quality criteria to various extents. Nevertheless, no structured analysis considering those criteria has been executed so far. It has rather been shown that popular role mining approaches like [43], [23] or [48] do not offer the guidance required to judge the correctness of role definitions and role models [18]. In practice, however, it is very likely that companies re-apply role mining techniques in order to ensure role model correctness after having employed a role mining approach for the initial role setup. Yet, none of the related work in the field focuses on quality criteria applied during role system maintenance. In [42], the authors investigated the usage of selected metrics like the Weighted Structural Complexity (WSC) for analyzing role system states. [20] described mechanisms for periodically evaluating a role systems quality but do not consider the scalability of their approach in large real-world scenarios. [18] recently proposed the integration of a distinct quality rating and role classification phase in their role optimization process model. However, they do not present an overview of available quality criteria and their application. As a result, the core challenge of measuring the current quality of a role model still remains unsolved.

In the following this paper improves the state of the art by firstly presenting a survey of role mining approaches and their considered quality metrics. A consecutive analysis of the quality criteria and their combined applicability aims at stimulating future research to integrate them into role mining in a modular manner. Based on company-specific quality criteria this would allow organizations to select the best fitting role mining approach in the first place and re-use it during periodic role model reviews.

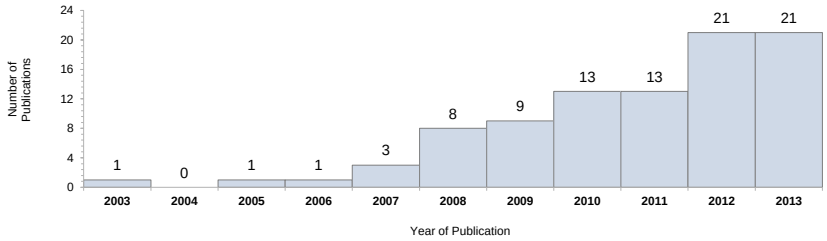

Fig. 2. Development of the research area

3 METHODOLOGY

Our methodology to answer the research questions presented in Section 1 is depicted in Figure 1. At first, we survey the research area and create a catalog of role mining approaches that serves as the basis for further analyses (step 1). For this literature survey, we follow the methodology proposed by [32]. We carried out a bibliographic database search including the ACM Digital Library[4], DBLP[5], IEEE Digital Library[6], and Google Scholar[7] using the keyword "role mining". To arrive at a complete catalog of role mining approaches, author and reference search for each publication was applied to identify previously undiscovered research. Finally, papers that do not present role mining techniques were removed from the catalog (e.g. [42]). Consecutively, we classify recent role mining publications according to the scheme presented by [19]. Additional clusters were added for role mining approaches that used new techniques. Note, that approaches that rely on more than one role mining technique were clustered based on the predominant technique being used. At this stage we completed our first contribution by extending the existing survey and highlighting the rapid increase of research within the last three years as well as the further diversification of role mining techniques used.

During step 2 of our methodology we answer RQ1 by investigating the usage of quality criteria in role mining algorithms. Note that we do not consider quality criteria known from other fields such as quality management in general as our work aims at stimulating research regarding role system maintenance in specific. General quality management indicators or processes can be incorporated in future research but are not required to rate the initial situation regarding quality criteria used during role system maintenance in the role mining community. Finally, we conduct an in-depth analysis of the identified quality criteria and their dependencies (step 3). To answer RQ2, our results are presented as an impact matrix, showing positive, negative, or no mutual influence.

[4] http://dl.acm.org/
[5] http://www.dblp.org/search/index.php
[6] http://www.computer.org/
[7] http://scholar.google.com/

4 DEVELOPMENT OF ROLE MINING RESEARCH

This section extends a survey on role mining research conducted by [19]. Figure 2 underlines the importance of an updated interpretation of role detection approaches due to the increase in researchers' attention during the last three years. Between 2011 and 2013, 55 papers related to role mining have been published, representing an increase of 141 percent compared to the number of publications from 2003 - 2010 (39).

While role mining in general consists of a pre-processing phase, a role detection phase, and a post-processing phase [19], the remainder of this survey focuses on the role detection phase as the core element of every role mining algorithm. During this phase, suitable roles are created based on an existing set of user-permission assignments (UPA). The algorithms can be grouped according to the underlying technique (see Table 1). While the first six techniques have been previously introduced in [19], we identified three additional techniques (Visual Role Mining, Boolean Matrix Decomposition, and Attribute-based Role Mining) being applied to solve the role mining problem for the first time. Additionally to the 21 algorithms published in [19], 26 out of the 55 approaches from 2011 to 2013 have been categorized, while the rest is related to either pre- or post-processing phase. Subsequently, each technique and representative publications are presented:

Technique	**Description**
Subset Enumeration	Calculates all possible intersections of permissions in the initial UPA
Clustering	Creates clusters with similar permissions that indicate role candidates
Graph Optimization	Convert an initial bipartite graph into a tripartite graph whereby middle vertices indicate role candidates
Frequent Permission Set Mining	Discovers permissions that frequently occur together in access control data
Formal Concept Analysis	Discovery of a hierarchical structure for representing the UPA-matrix through mathematical concepts
Heuristic Matrix Selection	Iteration through the rows of a UPA-matrix and selecting rows as roles based on heuristics
Visual Role Detection	Sorting and visualizing the UPA-matrix for detecting roles visually
Boolean Matrix Decomposition	Decomposition of the UPA-matrix into two consistent submatrices
Attribute-based Role Mining	Leveraging attributes to construct roles

Table 1. Overview of role mining techniques

Subset Enumeration aims to discover roles through creating all possible intersections of permission sets. Due to the exponential complexity of enumerating all possible subsets, algorithms such as [58] use heuristics for role candidate selection. Hence, role mining algorithms based on this technique strive to balance complexity and quality of

results. With 13 available algorithms, this is the most frequently used technique in role mining.

Clustering is a role mining approach that is directly derived from data mining [17]. Using a *UPA*-matrix as input, this technique searches for clusters with similar permissions. However, clustering approaches struggle with limitations such as requiring users or permissions to be only part of one single cluster [17]. To solve these challenges, several clustering variants exist. Examples include iterative application of the technique or reduction of the input [34],[28].

Graph Optimization uses bipartite graphs to represent the *UPA* [8]. As roles represent an intermediary between the two disjoint vertices of users and permissions, those approaches aim at converting the bipartite into a tripartite graph or a representation with sub-graphs. Roles are then represented by the middle vertex [8] or each necessary sub-graph [40].

Frequent Permission Set Mining has its roots in marketing analysis and algorithms of generic frequent set mining [1]. Initially used for the study of consumers' purchase behavior, it is applied by role mining algorithms in order to discover permissions that frequently occur together. The presumption is that permission sets that appear together can be interpresed as role candidates.

Formal Concept Analysis is a data mining technique similar to clustering, overcoming the limitation of only assigning entities to one group [42]. Related algorithms build a concept lattice from the *UPA*-matrix and reduce duplicate information. A concept lattice is a construct similar to a graph and represents roles and their connection in a partially ordered collection of clusters which again consist of permissions and users.

Heuristic Matrix Selection is similar to Subset Enumeration without the initial role set being based on intersections of permissions. Instead, it iterates through rows (or columns) and picks role candidates successively according to the highest number of given assignments (e.g permissions assigned to a user). Initially, each row/column is treated as a role. Subsequently, a cross-check for duplicate roles is conducted [3].

Visual Role Mining is a fairly new technique initially proposed in [10]. It reorders rows and columns of the input *UPA*-matrix in order to create clusters of adjoined permissions. Displayed to an administrator, the underlying assumption is that humans' cognitive capabilities and context knowledge are better suited to discover proper roles compared to purely algorithm-based approaches.

Boolean Matrix Decomposition is an approach that directly addresses the Role Mining Problem (RMP) introduced by [51]. This formal definition of role mining and targets at decomposing the boolean *UPA*-matrix into two separate matrices, a *UA*-matrix and a *PA*-matrix. By dividing the initial *UPA*-matrix into two consistent sub-matrices, the columns of the *UA*-matrix and the rows of the *PA*-matrix build up the set of roles.

Attribute-based Role Mining such as [16] are trying to incorporate business information through attributes into role mining. They rely on the assumption that additional semantic data is available and can be taken into account. Attribute-based approaches combine other techniques and enrich them with attribute-based mechanisms to arrive at an improved role set.

5 QUALITY-RELATED CRITERIA

5.1 Criteria

After their classification we examined role mining algorithms regarding their decision making processes of including certain role candidates in their final output. We argue that this central decision making provides well-suited indicators for quality management in RBAC. This assumption is based on the claims of several publications (e.g. [56], [63]) of outperforming competitive approaches in terms of the quality of generated roles.

A total of 23 different quality criteria can be identified. They either focus on the quality of the overall RBAC state, the quality of single roles, or both. At first we focus on RBAC state quality criteria. Secondly, we examine criteria that deal with the quality of an individual role (cf. Table 2). Note that for some criteria (e.g. *Exclude Unused Permissions*) additional input information is required. In the following, we present a detailed interpretation of the quality criteria and group them according to their focus.

Achieve Completeness

Completeness refers to the exact representation of the original access control state, i.e. the goal is to cover the initial UPA-matrix with the resulting set of roles. In contrast to most approaches (e.g. [60],[53],[3]), some techniques allow to deviate to a certain extent from the initial UPA-matrix based on a given threshold (the so called $\delta - RMP$ [51]) [5],[35]. Completeness therefore measures the quality of a RBAC state by measuring the degree to which a resulting role set represents the initial access control situation.

Reduce Number of Roles

Initially formulated in the RMP, the goal of having as few roles as possible is based on the assumption that complexity of RBAC is directly connected to the number of roles maintained. Thus, the number of roles in a given RBAC state is a quality criteria usable to rate the estimated administrative efforts to manage the role model. Depending on the size of a company in terms of its employees, user accounts, permissions and UPA, this measure can be normalized in order to allow for a comparison of role models in different organizations.

Decrease Role Set Similarity

Quality criteria related to Role Set Similarity measure the distance between two given sets of roles. They are mainly used for the measurement of the dissimilarity of RBAC states (e.g. in [62]) or the difference between a current RBAC state and a targeted state. In [29], for instance, the permission similarity is measured using the Euclidean Distance. Furthermore, the Jaccard Similarity is a popular metric used in a variety of approaches [40],[57],[5].

Minimize Users/Permissions per Role & Minimize/Maximize Roles per User/Permission

Quality criteria assigned to this category target at two main objectives. First, the definition of an upper bound per role limiting the amount of users assigned to one role. Second, the objective of finding as few as possible permissions that are placed in a role. This can, for instance, be applied in case an organization aims at defining a larger number of small roles for employees that exactly fit their specialist tasks. On the contrary, maximizing the number of users or permissions per role can be beneficial other scenarios, e.g. when organizations aim at defining a small set of large roles. Moreover, Minimizing/Maximizing the roles per element (user or permission) is applied in seven existing role mining approaches (e.g. see [34],[39]). This can be useful in case the overall role model complexity in terms of relationships among the role model elements should be minimized.

Fullfill Role Constraints

Role Constraints impose restrictions on the definition of roles. For instance, [38] consider Segregation of Duty (SOD) policies that entail mutually exclusive permissions in the RBAC state. Other policies, such as the four-eye-principle, that affect the user assignments of a role are also conceivable.

Reduce WSC

In contrast to most other quality criteria the so called Weighted Structural Complexity (WSC) is a widely-used heuristic to rate the complexity of a RBAC model. Originally introduced by [42], it applies weights to different optimization objectives. It can be seen as one of the most advanced measures that solely relies on the components of an RBAC state. It is usable for both, individual roles and role sets, and thus allows for a good comparability of RBAC states. As a result of its popularity, several existing role mining approaches are able to consider the WSC.

Optimize Matrix Sorting

Matrix sorting aims at covering an initial access control state by sorting the input UPA-matrix based on user accounts with similar permissions and permissions that are assigned a similar set of user accounts. [10] introduced the ADVISER and EXTRACT algorithms that generate a matrix representation of the initial UPA-matrix that clusters permissions and user accounts together. As a result, large areas covering initial UPA can be visually detected by a human role engineer.

Similarities & Redundancy

Well-known similarity metrics can be applied to the various elements of a RBAC state in order to measure its quality. [48] gives an overview of possible applications of the Jaccard Similarity in the context of role management. He discusses three similarity metrics that can be applied on the assignment types of a role (user, permission and

role hierarchy). They can further be used to compute the similarity of two role sets (cf. Decrease Role Set Similarity). Besides examining the similarity of assignment types of a role, similarity metrics are applied to attributes of role components. They can, for instance, be used to create a role set based on the location attribute of all user accounts. Distance measures are applied to identify redundant roles [5].

Increase Role Coverage

The Role Coverage is formally defined in [61] as the fraction of role-covered UPA by the initial UPA. Companies aim at achieving a high role coverage in order to foster the benefits of RBAC compared to other access control models. The implication of reducing administrative costs through RBAC is represented through this criterion.

Attribute-related Criteria

Attribute-related criteria evaluate the quality of a role based on its attributes or attributes of its components. Permission usage derived from access logs, for example, can be used to display the actual usage of privileges by employees. It offer insights into unused PA that can potentially be removed during the next refinement of a role [44]. Furthermore, restrictions on the composition of a role, e.g. by allowing only certain attributes of users in a role are possible [56].

5.2 Discussion

This work is motivated by the gap between the recent uprising of role mining and the practical need for periodic quality assessment of the resulting role models. The presented survey underlined the significant growth (141%) of published papers in the recent past. We have shown that every role mining approach relies on one or more quality criteria, mostly implicitly without providing a structured integration of quality management. In the following we present a short discussion of our quality-related findings from Figure 2.

Firstly, it can be seen that the main quality criterion in role mining is to arrive at an exact representation of existing access control states. This criterion is – to a varying extent – considered by all available approaches, except for [29] which derive the roles solely from access history logs. Secondly, Figure 2 shows that a large number of approaches focus on generating as few roles as possible (**Reduce Number of Roles**). Interestingly, as the WSC is a potential criterion which is able to represent this and other measures (by modifying its weight factors), recent approaches try to use this metric as a heuristic for producing high quality roles [58],[12],[59]. This can be interpreted as an indicator that research is already trying to integrate sophisticated measures – such as the WSC – into role mining.

Other interesting results are, that criteria with practical relevance up to now are only considered by few existing approaches (see Figure 2). Timestamps as an attribute of permissions are, for instance, only considered by one approach [41]. However, their integration into role mining seems promising as they heavily can influence role design. Sets of permissions activated together within a certain period of time can e.g. represent

Table 2. Quality Criteria (QC) in existing approaches

		QC	State			Individual Role							State + Individual Role											
			Achieve Completeness	Reduce Number of Roles	Decrease Role Set Similarity	Minimize Users per Role	Maximize Users per Role	Minimize Roles per User	Minimize Roles per Permission	Minimize Permissions per Role	Maximize Permissions per Role	Fulfill Role Constraints	Reduce WSC	Optimize Matrix Sorting	Decrease Permission Similarity	Reduce Role Redundancy	Increase User Similarity	Decrease Permission Attribute Similarity	Increase Role Coverage	Exclude Unused Permissions	Consider Timestamp	Consider Role Attributes	Consider User Attributes	Group by Attributes
Technique / Focus	**Paper**																							
Subset Enumeration	[5]		■	■			■								■	■								
	[6]		■	■			■				■		■											
	[28]		■	■			■				■								■					
	[34]		■	■				■			■								■					
	[41]		■	■																	■			
	[44]		■	■															■	■				
	[53]		■	■			■																	
	[51]		■	■															■					
	[50]		■	■				■			■								■					
	[52]		■	■																				
	[56]		■								■		■			■			■				■	
	[58]		■										■						■					
	[62]		■		■	■												■						
Clustering	[14]		■	■				■									■							
	[17]		■																					
	[15]		■															■				■		
	[31]		■							■														
	[46]		■	■																				
Graph Optimization	[8]		■	■																				
	[9]		■	■																				
	[11]		■	■		■				■														
	[22]		■													■				■				
	[26]		■	■		■																		
	[40]		■		■	■														■				
	[60]		■										■											
Frequent Permission Set Mining	[7]		■	■											■									
	[29]				■										■				■	■				
	[37]		■										■		■		■							
	[38]		■			■		■				■												
	[39]		■	■					■											■				
	[61]		■																■					
Formal Concept Analysis	[42]		■																					
	[54]		■	■																				
	[55]		■	■															■					
Heuristic Matrix Selection	[3]		■	■																				
	[4]		■			■				■		■												
	[27]		■	■			■				■													
Visual Role Mining	[10]		■	■										■					■					
	[12]		■	■									■	■					■					
Boolean Matrix Decomposition	[30]		■	■				■																
	[36]		■	■																				
	[35]		■	■																				
	[49]		■	■																				
	[59]		■	■									■											
Attribute-based Approaches	[16]		■	■																				■
	[25]		■																					■
	[33]		■					■																■

Uses criteria: ■ Yes ☐ No

good candidate permissions for a role. We argue that the low availability of timestamp information might be the main reason for its low acceptance in the community.

We furthermore noted that several quality criteria well-known in practice have not yet been included in any role mining approach at all. This includes criteria like the

- Maximum allowed number of roles in a role model
- Role usage
- Hierarchy restrictions

It seems straightforward to integrate a maximum threshold of roles to be found through automatic role discovery in order to ensure the maintainability of the whole role set. Intuitively, result sets can always be limited by just taking the desired number of roles after sorting them according to a predefined criterion. However, we argue that a dedicated parametrization of a role mining approach needs to be possible so that it considers this upper bound during the process of role generation. Furthermore, the usage of *UA* over a certain period of time (i.e. the activation of roles) can hint at outdated role definitions. Several approaches are able to take existing roles into consideration (e.g. [42]) but do not integrate usage data. Moreover, restrictions on the hierarchy of a RBAC state can represent one way to reduce complexity and increase the quality of either a single role or the whole RBAC state. In practice, deployed RBAC states feature unlimited depth, sometimes even resulting in hierarchy loops. Limiting the maximum allowed number of parent or child roles of a role or the maximal hierarchy depth can ease administrative staff's understanding of the overall role model. Note that [15] already considers RBAC hierarchies. However, they only introduce an overall hierarchy depth of two and indicate the possibility to extend their probabilistic approach with more layers. Several post-processing approaches already outline the need for an inspection of the role hierarchy (e.g. [24],[48]). However, they only focus on removing duplicate hierarchy depiction and finding the minimal set of hierarchical assignments, not on cardinality restrictions.

5.3 RELATIONSHIPS BETWEEN QUALITY CRITERIA

After having presented the set of quality criteria currently used in role mining, in this section we analyze the relationships between different quality criteria in order to answer RQ2. By showing how the various quality categories affect each other (see Figure 3), we point at potential combinations that can be applied during strategic role model maintenance. Knowledge about quality criteria and their mutual influence can support companies during the selection of the best fitting approach in a given scenario. It is of major importance to consider strategic role maintenance efforts before designing the initial role model. A company might select a certain role mining approach (that even may require a higher initial role definition effort) as its future maintenance is expected to be significantly lower in the long run. Up to now, such a qualified role mining selection is not possible.

Figure 3 illustrates mutual dependencies between quality criteria. A positive influence implies that fulfilling one criteria impacts another criteria in a way that it can be easier or more efficiently achieved (and vice versa). Negative influence means that focusing on one criteria impairs the fulfillment of the other criteria. A white background

Fig. 3. Dependencies between quality criteria

	Achieve Completeness	Reduce Number of Roles	Decrease Role Set Similarity	Minimize Users per Role	Maximize Users per Role	Minimize Roles per User	Minimize Roles per Permission	Minimize Permissions per Role	Maximize Permissions per Role	Fulfill Role Constraints	Reduce WSC	Optimize Matrix Sorting	Decrease Permission Similarity	Reduce Role Redundancy	Increase User Similarity	Decrease Permission Attribute Similarity	Increase Role Coverage	Exclude Unused Permissions	Consider Timestamp	Consider Role Attributes	Consider User Attributes	Group by Attributes
Achieve Completeness																						
Reduce Number of Roles	-																					
Decrease Role Set Similarity	-	-																				
Minimize Users per Role	-	-	0																			
Maximize Users per Role	-	+	0	-																		
Minimize Roles per User	-	-	+	-	+																	
Minimize Roles per Permission	-	-	+	0	0	0																
Minimize Permissions per Role	-	-	0	0	0	0	-															
Maximize Permissions per Role	-	+	0	0	0	0	+	-														
Fulfill Role Constraints	-	-	+	-	-	-	-	-	-													
Reduce WSC	-	+	0	+	+	+	+	+	+	-												
Optimize Matrix Sorting	-	0	+	0	0	0	0	0	0	0	0											
Decrease Permission Similarity	-	-	+	0	0	0	0	0	0	0	0	+										
Reduce Role Redundancy	-	-	+	0	0	0	0	0	0	0	0	+	+									
Increase User Similarity	-	-	+	0	0	0	0	0	0	0	0	+	0	+								
Decrease Permission Attribute Similarity	-	-	+	0	-	0	0	0	-	0	0	0	+	+	+							
Increase Role Coverage	-	-	0	-	+	-	-	-	+	-	+	-	0	0	0	0						
Exclude Unused Permissions	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-					
Consider Timestamp	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0				
Consider Role Attributes	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0			
Consider User Attributes	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	+		
Group by Attributes	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	+	0	

Impact: Positive None/Situation-dependent Negative

is used, if no definite statement can be made. For our analysis, we focus on direct implications between two criteria and evaluate those effects if possible. While we are aware of potential differences in the impact intensity, the goal of our research is not to examine the degree of influences. It rather is to provide a first overview of dependencies and stimulate further research in the area. We argue that without the knowledge of the currently applied quality criteria, a structured integration – above all considering criteria combinations – into role mining approaches is hardly possible.

The matrix presented in Figure 3 reveals that in order to **Achieve Completeness**, negative impacts on all other except attribute-based criteria are the consequence. Intuitively, an algorithm that aims at covering the input UPA matrix with roles leaves few possibilities to optimize other criteria such as maximizing the number of users per role. A complete coverage of an input UPA state leads to an increased role count and a decreased number of users per role.

Similarly, **Reduce the Number of Roles** negatively impacts an optimization towards most other quality criteria. Exceptions are the maximization of users and permissions per role as reducing the number of roles inherently entails an increase of users and

permissions per role. Likewise, a small number of roles has a positive influence (e.g. on the WSC) since the number of roles is given less weight.

Given the goal to **Decrease Role Set Similarity** has several side-effects on the fulfillment of other criteria. As users and permissions two role sets have in common are used to measure similarity, minimizing the number of roles per user leads to more distinct roles. In general, a decreasing Role Set Similarity positively impacts the decrease of other similarity-related criteria as its calculation is commonly based on the similarity of role components. Among negative impacts is its influence on the reduction of the total role number and the completeness to be achieved. Both originate from the fact that regulations on the similarity of roles lead to an increased need of new role definitions for representing the access control.

Techniques that **Minimize Users per Role** positively impact the overall WSC. A decrease of UA, e.g., leads to a lower general WSC. On the contrary, negative dependencies exist between minimizing the number of users per role and minimizing the number of roles per user. Intuitively, fewer users per role require the assignment of more roles to a single user in order to arrive at the same number of privileges.

The goal of maximizing the number of users per role (**Maximize Users per Role**) impairs fulfilling other criteria also related to role configuration and role size. A positive impact on finding a minimum role set and role coverage can be inferred as maximizing the number of users per role leads to fewer and larger roles that cover a significant portion of the UPA-matrix.

Techniques that **Minimize Roles per User** positively influence the overall WSC in case URA are considered in the WSC calculation. Likewise, minimizing the number of roles requires more users to be assigned to a role. Yet, considering this criterion leads to negative effects on role coverage and role constraints. Fewer roles promote the use of direct UPA and lead to a lower role coverage. Role constraints are negatively impacted as role design becomes more difficult if only a few roles per user are allowed.

The goal of minimizing the number of roles per permission (**Minimize Roles per Permission**) affects other criteria in the same way. It leads to a reduced number of role-permission assignments which has a positive impact on the overall WSC. Negative impacts can be expected for role constraints due to increased role definition complexity and for role coverage due to a larger number of direct UPA.

Dependencies for **Minimize Permissions per Role** and **Maximize Permissions per Role** are similar to the respective minimization and maximization criteria for users per role.

Role mining techniques that aim to **Fulfill Role Constraints** in general struggle with optimizing most other criteria except for a decreased Role Set Similarity. The positive impact in this case stems from the fact that role definition constraints are often referring to SoD policies which in return are likely to produce a disjunctive – and thus dissimilar – set of roles.

Trying to **Reduce WSC** positively affects several other criteria, in particular those that aim at decreasing the number of role component assignments. A negative impact can be observed when using this criterion in combination with either the achievement of completeness or role constraints. Regarding completeness an increased number of roles to cover the input UPA negatively influences the WSC. Furthermore, as the percentage

of *UPA* covered by roles has a significant impact on lowering the WSC, using role constraints in general increases the WSC.

As **Optimize Matrix Sorting** aims at creating clusters of similar users and roles, it is likely to have a positive impact on other similarity-related criteria. When sorted next to each other in a visualized *UPA*-matrix, a human role engineer is supported during tasks like merging or separating roles according to their similarity. Using visualization filters can allow the role engineer to display information according to similarity (e.g. similar job positions of employees) and model roles accordingly.

Decrease Role Permission Similarity-related techniques that focus on creating roles with distinct permissions inherently have a positive impact on reducing role redundancy as they decrease the number of overlapping roles. Similarly, they positively impact attribute similarity as they generate attribute-based roles. Commonly different attribute values (e.g. location or function of employees) lead to roles that are likely to bundle distinct sets of permissions.

As mentioned before, the goal to **Decrease Role Redundancy** positively affects all similarity-related criteria. Intuitively, minimizing redundancy has a negative impact on criteria such as completeness and finding the minimal number of roles. Completeness typically generates a large number of roles with redundant permissions in order to cover an initial *UPA*-matrix to a certain extent. Likewise, trying to reduce the overall number of roles inherently leads to larger roles with overlapping permissions.

Increase Role Coverage negatively affects the accomplishment of most other quality criteria. Exceptions include the maximization of users and permissions per role which are direct consequences of trying to maximize role coverage. Additionally, the fact that all *UPA* are covered by roles positively impacts the WSC as the usage of direct *UPA* increasing complexity is minimized.

Techniques aiming at the **Decrease of Permission Attribute Similarity** argue that permissions with similar attributes may be redundant and thus should not be included into one role. This implies a positive impact on Role Permission Similarity as roles are more likely to have distinct permissions. Consequently, the positive impact on Role Set Similarity stems from the fact that Role Set Similarity is often measured on the basis of the similarity of the permissions of a specific role [63],[29].

The remaining **Attribute-related Criteria**, namely **Exclude Unused Permissions**, **Consider Timestamp**, **Consider Role Attributes**, **Consider User Attributes**, and **Group by Attributes** share that they impose restrictions on the definition of the role catalog and thus have a negative impact on the fulfillment of most other criteria. For example, practical projects with partners show that over a certain period of time not activated permissions by the users of a role should not be included. However, the affected *UPA* still exist and therefore will be taken into account by role mining. An a priori cleansing of the input data through a structured role optimization process [18] can aid in avoiding such problems, but has not directly been placed into role detection mechanisms.

6 CONCLUSIONS AND FUTURE WORK

Role-based Access Control as the de facto standard for managing access privileges in organizations struggles with the dynamic evolution of role models over time. As a result the quality of RBAC states initially modeled using role mining techniques decreases over time. In order to address this challenge, role mining mechanisms applied during role system maintenance need to be extended in order to integrate a dedicated quality management stage for rating and improving a role system state on the basis of company-specific quality criteria. In this paper we presented three contributions in that respect. We firstly provided a survey giving an overview of current role mining approaches. The significant increase of research activity and the growing number of applied techniques for generating roles during the last three years underlines the relevance and diversification of the area. By extracting criteria that are dedicated to improving role quality from currently available role mining approaches we were able to answer RQ1 in Section 4. In Section 5.3 we then analyzed the identified quality criteria and their mutual dependencies in order to answer RQ2. We have shown on which quality criteria current role mining approaches rely and revealed a number of practically relevant but yet untreated criteria in research. The results highlight the need for a structured integration of quality mechanisms into role mining in order to allow for an improved selection of role mining approaches in a given scenario based on company-specific quality criteria. Up to now, our findings are restricted to the field of role mining exclusively. Expanding the scope towards areas like Quality Management or Data Mining in general could yield additional criteria. In future work we are thus going to investigate the promising concept of integrating Quality Management Frameworks such as the EFQM Excellence Model [8].

Acknowledgments. The research leading to these results was supported by the "Bavarian State Ministry of Education, Science and the Arts" as part of the FORSEC research association. This work would not have been possible without our student Christian Wawarta.

References

1. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. In: SIGMOD Record. vol. 22, pp. 207–216. ACM (1993)
2. Basel Comittee on Banking Supervisions: Basel III: Int. framework for liquidity risk measurement, standards and monitoring (2010)
3. Blundo, C., Cimato, S.: A simple role mining algorithm. In: Proc. of the 2010 Symp. on Applied Computing (SAC). ACM (2010)
4. Blundo, C., Cimato, S.: Constrained role mining. In: Proc. of the 8th Int. Workshop on Security and Trust Management (STM), pp. 289–304. Springer (2013)
5. Chu, V.W., Wong, R.K., Chi, C.H.: Over-fitting and error detection for online role mining. Int. Journal of Web Services Research 9(4), 1–23 (2012)
6. Colantonio, A., Di Pietro, R., Ocello, A.: A cost-driven approach to role engineering. In: Proc. of the 2008 Symp. on Applied Computing (SAC). ACM (2008)

[8] http://www.efqm.org/

7. Colantonio, A., Di Pietro, R., Ocello, A.: Leveraging lattices to improve role mining. In: Proc. of the IFIP TC-11 23rd Int. Information Security Conf. (SEC). Springer (2008)
8. Colantonio, A., Di Pietro, R., Ocello, A., Verde, N.V.: A probabilistic bound on the basic role mining problem and its applications. In: Proc. of the 24th IFIP TC-11 Int. Information Security Conf. (SEC), pp. 376–386. Springer (2009)
9. Colantonio, A., Di Pietro, R., Ocello, A., Verde, N.V.: Taming role mining complexity in rbac. Computers & Security 29(5), 548–564 (2010)
10. Colantonio, A., Di Pietro, R., Ocello, A., Verde, N.V.: Visual role mining: A picture is worth a thousand roles. IEEE Transactions on Knowledge and Data Engineering 24(6), 1120–1133 (2012)
11. Ene, A., Horne, W., Milosavljevic, N., Rao, P., Schreiber, R., Tarjan, R.E.: Fast exact and heuristic methods for role minimization problems. In: Proc. of the 13th Symp. on Access Control Models and Technologies (SACMAT). ACM (2008)
12. Eucharista, A.and Haribaskar, K.: Visual elicitation of roles: using a hybrid approach. Oriental Journal of Computer Science & Technology 6(1), 103–110 (2013)
13. European Union: General data protection regulation (2012)
14. Frank, M., Basin, D., Buhmann, J.M.: A class of probabilistic models for role engineering. In: Proc. of the 15th ACM Conf. on Computer and Communications Security (CCS). ACM (2008)
15. Frank, M., Buhman, J.M., Basin, D.: Role mining with probabilistic models. ACM Transactions on Information and System Security (TISSEC) 15(4), 15:1–15:28 (2013)
16. Frank, M., Streich, A.P., Basin, D., Buhmann, J.M.: A probabilistic approach to hybrid role mining. In: Proc. of the 16th ACM Conf. on Computer and Communications Security (CCS). pp. 101–111. ACM (2009)
17. Frank, M., Streich, A.P., Basin, D., Buhmann, J.M.: Multi-assignment clustering for boolean data. Journal of Machine Learning Research 13(1), 459–489 (2012)
18. Fuchs, L., Kunz, M., Pernul, G.: Role model optimization for secure role-based identity management. In: Proc. of the 22nd European Conf. on Information Systems (ECIS) (2014)
19. Fuchs, L., Meier, S.: The role mining process model - underlining the need for a comprehensive research perspective. In: Proc. of the 6th Int. Conf. on Availability, Reliability and Security (ARES). IEEE (2011)
20. Fuchs, L., Müller, C.: Automating periodic role-checks: A tool-based approach. In: Business Services: Konzepte, Technologien, Anwendungen: 9. Int. Tagung Wirtschaftsinformatik (WI), vol. 246. OCG, Wien (2009)
21. Fuchs, L., Pernul, G., Sandhu, R.: Roles in information security–a survey and classification of the research area. Computers & Security 30(8), 748–769 (2011)
22. Gal-Oz, N., Gonen, Y., Yahalom, R., Gudes, E., Rozenberg, B., Shmueli, E.: Mining roles from web application usage patterns. In: Proc. of the 8th Int. Conf. on Trust, Privacy and Security in Digital Business (TrustBus), Lecture Notes in Computer Science, vol. 6863, pp. 125–137. Springe (2011)
23. Giblin, C., Graf, M., Karjoth, G., Wespi, A., Molloy, I., Lobo, J., Calo, S.B.: Towards an integrated approach to role engineering. In: SafeConfig. pp. 63–70. ACM (2010)
24. Guo, Q., Vaidya, J., Atluri, V.: The role hierarchy mining problem: Discovery of optimal role hierarchies. In: Proc. of the 24th Computer Security Applications Conf. (ACSAC). IEEE (2008)
25. Han, D.j., Zhuo, H.k., Xia, L.t., Li, L.: Permission and role automatic assigning of user in role-based access control. Journal of Central South University 19, 1049–1056 (2012)
26. Hingankar, M., Sural, S.: Towards role mining with restricted user-role assignment. In: 2nd Int. Conf. on Wireless Communication, Vehicular Technology, Information Theory and Aerospace Electronic Systems Technology (Wireless VITAE) (Feb 2011)

27. Huang, C., Sun, J.l., Wang, X.y., Si, Y.j.: Minimal role mining method for web service composition. Journal of Zhejiang University SCIENCE C 11(5), 328–339 (2010)
28. Huang, H., Shang, F., Zhang, J.: Approximation algorithms for minimizing the number of roles and administrative assignments in rbac. In: Proc. of the 36th Annual Computer Software and Applications Conf. Workshops (COMPSAC). IEEE (2012)
29. Jafari, M., Chinaei, A., Barker, K., Fathian, M.: Role mining in access history logs. Journal of Information Assurance and Security 38 (2009)
30. John, J., Sural, S., Atluri, V., Vaidya, J.: Role mining under role-usage cardinality constraint. In: Proc. of the 27th Information Security and Privacy Research Conf. (SEC), IFIP Advances in Information and Communication Technology, vol. 376, pp. 150–161. Springer (2012)
31. Kumar, R., Sural, S., Gupta, A.: Mining rbac roles under cardinality constraint. In: Proc. of the 6th Int. Conf. on Information Systems Security (ICISS), pp. 171–185. Springer (2011)
32. Levy, Y., Ellis, T.J.: A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research. Informing Science Journal 9, 181–212 (2006)
33. Li, R., Wang, W., Ma, X., Gu, X., Wen, K.: Mining roles using attributes of permissions. Int. Journal of Innovative Computing, Information and Control 8(11), 7909–7924 (2012)
34. Lu, H., Hong, Y., Yang, Y., Duan, L., Badar, N.: Towards user-oriented rbac model. In: Proc. of the 27th Conf. on Data and Applications Security and Privacy XXVII (DBSec), Lecture Notes in Computer Science, vol. 7964, pp. 81–96. Springer (2013)
35. Lu, H., Vaidya, J., Atluri, V.: Optimal boolean matrix decomposition: Application to role engineering. In: Proc. of the 24th IEEE Int. Conf. on Data Engineering (ICDE). IEEE (2008)
36. Lu, H., Vaidya, J., Atluri, V., Hong, Y.: Constraint-aware role mining via extended boolean matrix decomposition. IEEE Transactions on Dependable and Secure Computing (TDSC) 9(5), 655–669 (2012)
37. Ma, X., Li, R., Lu, Z.: Role mining based on weights. In: Proc. of the 15th Symp. on Access Control Models and Technologies (SACMAT). ACM (2010)
38. Ma, X., Li, R., Lu, Z., Wang, W.: Mining constraints in role-based access control. Mathematical and Computer Modelling 55(1), 87–96 (2012)
39. Ma, X., Tian, Y., Zhao, L., Li, R.: Mining role based on ranks. Int. Journal of Research and Surveys - ICIC Express Letters. Part B, Applications 4(2), 319–326 (2013)
40. Mandala, S., Vukovic, M., Laredo, J., Ruan, Y., Hernandez, M.: Hybrid role mining for security service solution. In: Proc. of the 9th Int. Conf. on Services Computing (SCC). IEEE (2012)
41. Mitra, B., Sural, S., Atluri, V., Vaidya, J.: Toward mining of temporal roles. In: Proc. of the 27th Conf. on Data and Applications Security and Privacy XXVII (DBSec), Lecture Notes in Computer Science, vol. 7964, pp. 65–80. Springer (2013)
42. Molloy, I., Chen, H., Li, T., Wang, Q., Li, N., Bertino, E., Calo, S., Lobo, J.: Mining roles with semantic meanings. In: Proc. of the 13th Symp. on Access Control Models and Technologies (SACMAT). ACM (2008)
43. Molloy, I., Chen, H., Li, T., Wang, Q., Li, N., Bertino, E., Calo, S., Lobo, J.: Mining roles with multiple objectives. In: ACM Transactions on Information and System Security (TISSEC). ACM (2010)
44. Molloy, I., Park, Y., Chari, S.: Generative models for access control policies: Applications to role mining over logs with attribution. In: Proc. of the 17th Symp. on Access Control Models and Technologies (SACMAT). ACM (2012)
45. Sandhu, R.S., Coyne, E.J., Feinstein, H.L., Youman, C.E.: Role-based access control models. Computer 29(2), 38–47 (1996)
46. Schlegelmilch, J., Steffens, U.: Role mining with orca. In: Proc. of the 10th Symp. on Access Control Models and Technologies (SACMAT). ACM (2005)
47. SOX: Sarbanes-oxley act of 2002, pl 107-204, 116 stat 745 (July 2002)

48. Takabi, H., Joshi, J.B.: Stateminer: An efficient similarity-based approach for optimal mining of role hierarchy. In: Proc. of the 15th Symp. on Access Control Models and Technologies (SACMAT). ACM (2010)
49. Uzun, E., Atluri, V., Lu, H., Vaidya, J.: An optimization model for the extended role mining problem. In: Proc. of the 25th Conf. on Data and Applications Security and Privacy XXV (DBSec), pp. 76–89. Springer (2011)
50. Vaidya, J., Atluri, V., Warner, J., Guo, Q.: Role engineering via prioritized subset enumeration. IEEE Transactions on Dependable and Secure Computing (TDSC) 7(3), 300–314 (July 2010)
51. Vaidya, J., Atluri, V., Guo, Q.: The role mining problem: finding a minimal descriptive set of roles. In: Proc. of the 12th Symp. on Access Control models and Technologies (SACMAT). ACM (2007)
52. Vaidya, J., Atluri, V., Guo, Q.: The role mining problem: A formal perspective. ACM Transactions on Information and System Security (TISSEC) 13(3), 27 (2010)
53. Vaidya, J., Atluri, V., Warner, J.: Roleminer: Mining roles using subset enumeration. In: Proc. of the 13th ACM Conf. on Computer and Communications Security (CCS). ACM (2006)
54. Wang, J., Zeng, C., He, C., Hong, L., Zhou, L., Wong, R.K., Tian, J.: Context-aware role mining for mobile service recommendation. In: Proc. of the 27th Annual Symp. on Applied Computing (SAC). ACM (2012)
55. Wong, R.K., Chu, V.W., Hao, T., Wang, J.: Context-aware service recommendation for moving connected devices. In: Proc. of the Int. Conf. on Connected Vehicles and Expo (ICCVE) (Dec 2012)
56. Xu, Z., Stoller, S.D.: Algorithms for mining meaningful roles. In: Proc. of the 17th Symp. on Access Control Models and Technologies (SACMAT). ACM (2012)
57. Xu, Z., Stoller, S.D.: Mining attribute-based access control policies from rbac policies. In: Proc. of the 10th Int. Conf. and Expo on Emerging Technologies for a Smarter World (CEWIT). IEEE (2013)
58. Xu, Z., Stoller, S.D.: Mining parameterized role-based policies. In: Proc. of the 3d ACM Conf. on Data and Application Security and Privacy (CODASPY). ACM (2013)
59. Ye, W., Li, R., Li, H.: Role mining using boolean matrix decomposition with hierarchy. In: Proc. of 12th IEEE Int. Conf. on Trust, Security and Privacy in Computing and Communications (TrustCom). IEEE (2013)
60. Zhang, D., Ramamohanarao, K., Ebringer, T.: Role engineering using graph optimisation. In: Proc. of the 12th Symp. on Access Control Models and Technologies (SACMAT). ACM (2007)
61. Zhang, D., Ramamohanarao, K., Ebringer, T., Yann, T.: Permission set mining: Discovering practical and useful roles. In: Proc. of the 24th Annual Computer Security Applications Conf. (ACSAC). IEEE (2008)
62. Zhang, W., Chen, Y., Gunter, C., Liebovitz, D., Malin, B.: Evolving role definitions through permission invocation patterns. In: Proc. of the 18th Symp. on Access Control Models and Technologies (SACMAT). ACM (2013)
63. Zhang, X., Han, W., Fang, Z., Yin, Y., Mustafa, H.: Role mining algorithm evaluation and improvement in large volume android applications. In: Proc. of the 1st Int. Workshop on Security in Embedded Systems and Smartphones (SESP). ACM (2013)
64. Zhu, H., Zhou, M.: Roles in information systems: A survey. IEEE Transactions on Systems, Man, and Cybernetics (SMC) 38(3), 377–396 (2008)